HYPEBEAST

Data Analytics Report

Report Summary

The datasets given are the sales and traffic access of customers in 2020 as well as products information and customer information. This report is made not only to answers the three management questions, but also to reveal extra insights base on the datasets.

To cut a long story short , based on the analytics, it can be concluded that :

1. The website traffic has little impact on the revenue.
2. Top 5 highest revenue products are not products that have highest pageviews.
3. There are 4 main customer segments :
   - Junior customers from age 11-30 who don't purchase much (191 customers)
   - Junior customers from age 11-30 who purchase a lot (177 customers)
   - Senior customers from age 30 – 56 who don't purchase much (208 customers)
   - Seniors customers from age 30 – 56 who purchase a lot. (194 customers)

4. Extras :

- German customers are the most revenue generated group

- Demographic factors such as age , gender and country have little impact on their spending.

- Newsletter subscribe customers contributes approximately 20% more revenue than who do not

- Italy and the UK customer has the largest percentage of 6 year loyalty age in total customers

Further explainations will be provided in the Findings

Contents

2 Findings

1 The website traffic has little impact on the revenue

Since our pageviews ,uniquepageviews and users access are not significantly identical throughout the period, we can use any of the three indicators to measure traffic. (Appendix 1)

Also , the total revenue between In Store and Online Store are not significantly identical throughout the period , it is unnecessary to separetly analyse them in the total revenue. (Appendix 2)

Hence , we can use the general total revenue for analysing.

A glance at the plots of total revenue to see the similarity between In Store and Online Channel, further statistical descriptions are provided in the Appendix 2 to see that similarity between In Store and Online Channel.
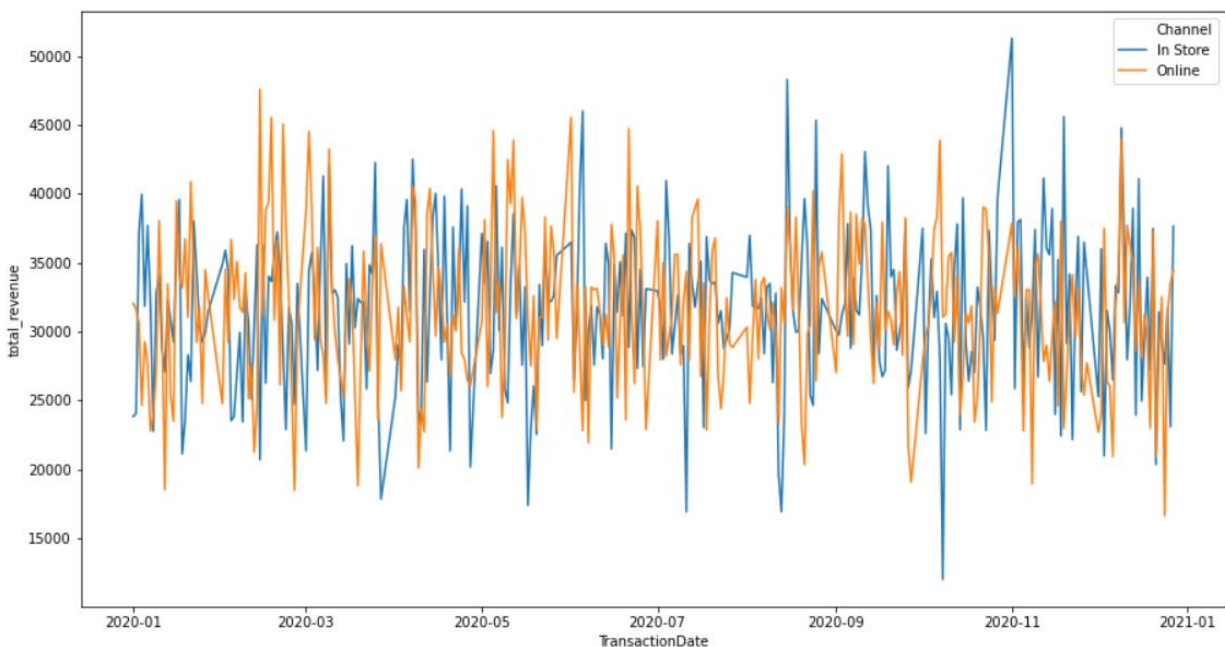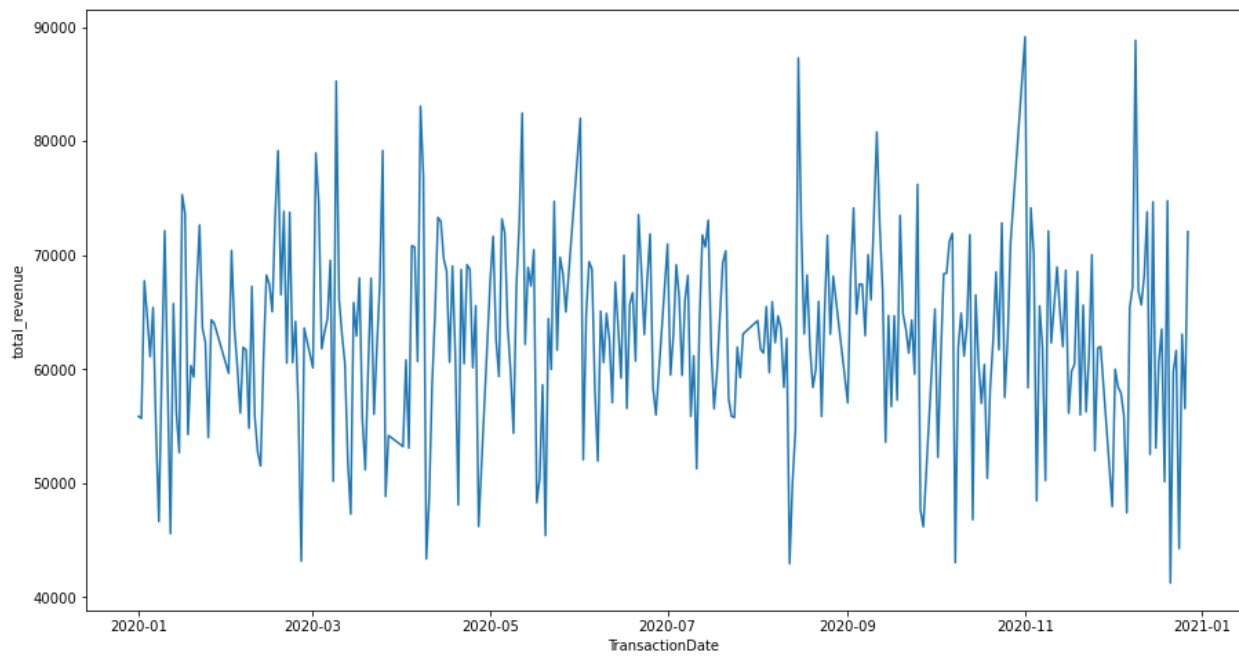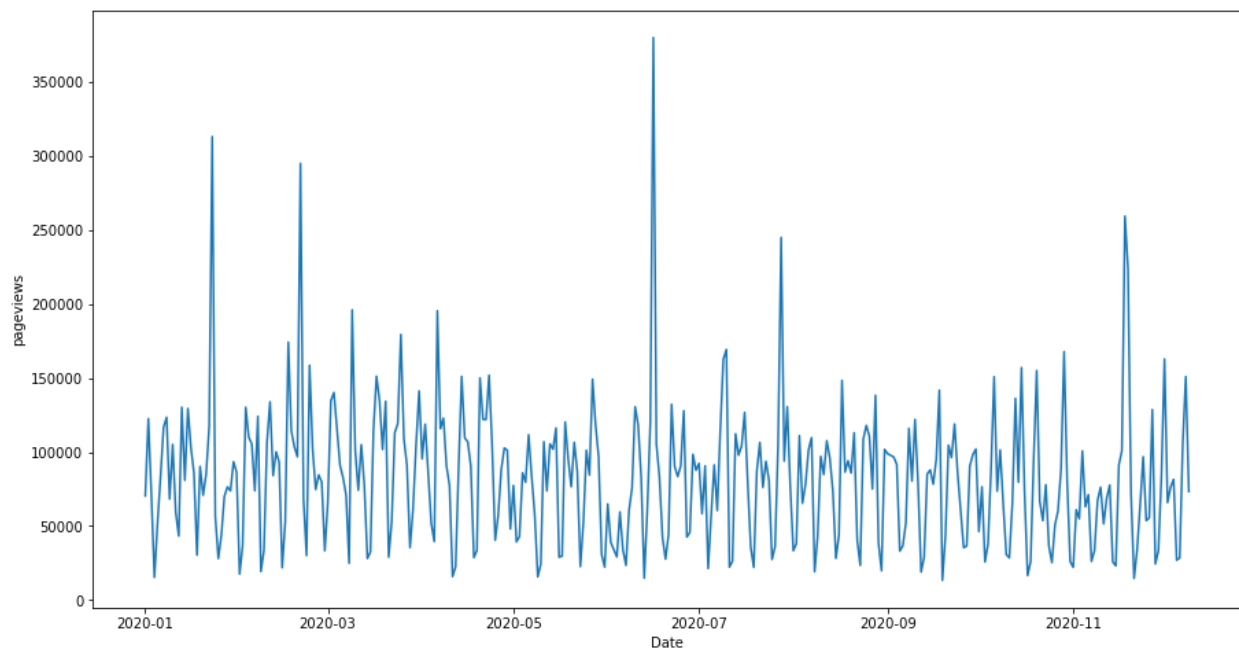

Figure 1 : Total Revenue from 2021-01 and 2020-12 of both Channel

Finally, the comparison between daily Revenue and daily Pageviews in 2020.

Total Revenue in 2020 by Date

Total Pageviews in 2020 by Date

In addition , There are not correlations, or relationships between daily Pageviews and daily revenue when show them in a scatterplot.
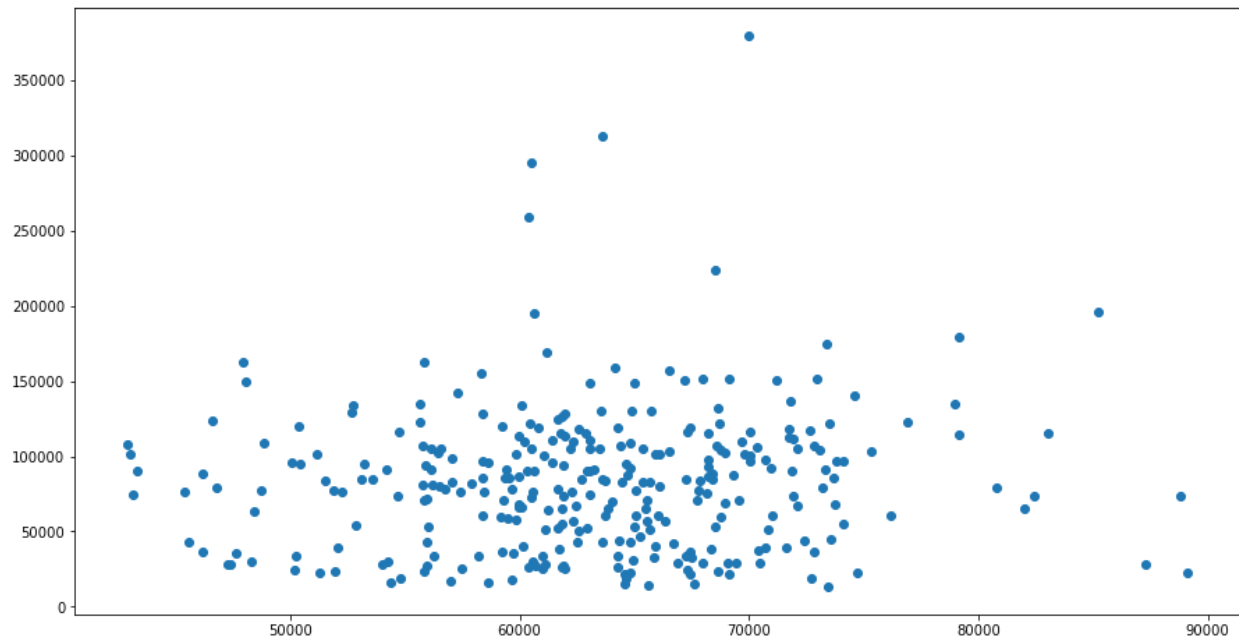
The correlation between the pageviews and total revenue are also measured, which is 0.075, very small and close to 0, which indicates there are no significant correlation between the two variables. As a result, we dont see any patterns between the two variables in the chart.

Notes :

- Correlation > 1 => There are positive correlation between 2 variables
- Correlation < 1 => There are negative correlation between 2 variables
- Correlation = 0 => There are no correlation between 2 variables

2. Top Products that have good total revenue and pageviews

These are the five products that has the highest revenue , but very normal pageviews

	Brand	Product	total_revenue	pageviews	Ranking	Traffic_Ranking
4885	II^(xPdB:S`#irqz	hypsrview0.321288570724117	11556	587.0	Cash Cow	Medium
9874	Relp\+KJ?D,cWw0P	reeelease0.452821711209563	9205	281.8	Cash Cow	Low
306	Ac8IJsKH,4xtY.Tk	audE-info0.884915261087885	8428	1205.6	Cash Cow	High
3195	DiCo4(99zZ<nkafj	kinv-news0.29905739542661	7980	766.4	Cash Cow	High
12575	YMbpE$ev3qMx-h*E	ymc_orson0.319251813809483	7794	223.2	Cash Cow	Low

These are the five products that is amongst the highest revenue and highest pageviews

	Brand	Product	total_revenue	pageviews	Ranking	Traffic_Ranking
10155	Sa?9zXUH5iJbuE'S	samLement0.78565851365938	7518	1590.0	Cash Cow	Most Popular
5655	Ka>n{\}.q.~P&*Ao	takgeveal0.557457939403364	7049	8596.2	Cash Cow	Most Popular
7235	MouDd/fn,XEARhBp	stupponyo0.211329698054268	6954	7190.6	Cash Cow	Most Popular
9488	PoAqnL=>P9Qb"ZUa	poryn-9920.3097516169666229	6815	10795.6	Cash Cow	Most Popular
9780	Relp\+KJ?D,cWw0P	bbcF-date0.977769785530841	6815	2049.6	Cash Cow	Most Popular

These are the five products that is amongst the highest revenue but lowest pageviews

	Brand	Product	total_revenue	pageviews	Ranking	Traffic_Ranking
9874	Relp\+KJ?D,cWw0P	reeelease0.452821711209563	9205	281.8	Cash Cow	Low
12575	YMbpE$ev3qMx-h*E	ymc_orson0.319251813809483	7794	223.2	Cash Cow	Low
1407	Ap8r`F[}sfP ->SZ	ubeYanada0.100113795307302	7552	121.0	Cash Cow	Low
3626	Exa~e4sZi*|fpD<^	ausDition0.316805159126848	7232	161.4	Cash Cow	Low
6900	Mi+QB'FqF;cGT)Y'	theftream0.650769488105747	7000	350.2	Cash Cow	Low

These are the five products that is amongst the highest pageviews but lowest revenue

	Brand	Product	total_revenue	pageviews	Ranking	Traffic_Ranking
6582	MaC0+mG3L{f>6+l	umbB-date0.0113296400042067	741	2129.0	Low	Most Popular
11706	Tr1)L/A]=of['Qhn	chaPtopia0.305759133071658	741	1906.0	Low	Most Popular
13004	ad(;%f6iD'}9EHD[	whil-info0.511259246630254	741	1737.6	Low	Most Popular
1397	ApIFulqmT[82a2/E	youTcture0.944314892191284	740	3566.4	Low	Most Popular
8553	NiPea$n|Eu@<@>'L	niku5-3000.815662316070307	740	2231.6	Low	Most Popular

Maybe the dataset was evenly distributed at the start, relationship matrix between pageviews and revenue is very evenly distributed, which can be examined in the notebook.

		Ranking
Ranking	Traffic_Ranking	
Cash Cow	High	840
	Low	852
	Medium	864
	Most Popular	833
High	High	834
	Low	832
	Medium	845
	Most Popular	877
Low	High	857
	Low	867
	Medium	827
	Most Popular	832
Medium	High	857
	Low	821
	Medium	866
	Most Popular	838

Notes : The results and the entire Ranking analysis are calculated based on Interquartile Range methods, which is in the notebook included in the appendix.

3.There are 4 Customer Segments

Since Demographic Factors such as Age , gender, country do not contribute significantly in total revenue (see part 4 Extras and Appendix 3). Factors such as Loyalty , Monetary and Youth are used to perform Customer Segmentation.

- Loyalty : Loyalty is the difference between today and date joined of customers. Although this factor has little impact in customer spending, it is included in the algorithm.
- Monetary : Customer's Monetary is the ranking of their revenue compare to other customers . Monetary varies from 1 to 4 , with 1 is lowest and 4 is highest.
- Youth : Customer's Youth is their age. Youth varies from 1 to 4 , with 1 is lowest and 4 is highest.

The value from 1 to 4 in each factor is determined using Interquartile Range.

After performing analysis using Machine Learning (KMeans algorithm). It can be seen that there are 4 main customer segments amongst our customers, each group is equally and evenly distributed.

- Junior customers from age 11-30 who don't purchase much (191 customers)

- Junior customers from age 11-30 who purchase a lot (177 customers)

- Senior customers from age 30 – 56 who don't purchase much (208 customers)

- Seniors customers from age 30 – 56 who purchase a lot. (194 customers)

The entire analytic process is included in the notebook at the end of this report for further inspection.

4. Extras

4.1 German customers are the most revenue group

Based on the data given, customer demographic factors such as age , gender are very evenly and equally distributed, as a result, it is ambiguous to segment customers in terms of age and gender.

However, it can be seen that more than half of the customers are German , the second largest group is French, followed by Italy. UK people only account for approximately 10% of our customers.

```
GER - Germany          321
FR - France            256
IT - Italy             128
UK - United Kingdom     65
```

Based on the country ratio in the customers dataset, German customers account for the largest revenue, followed by French.
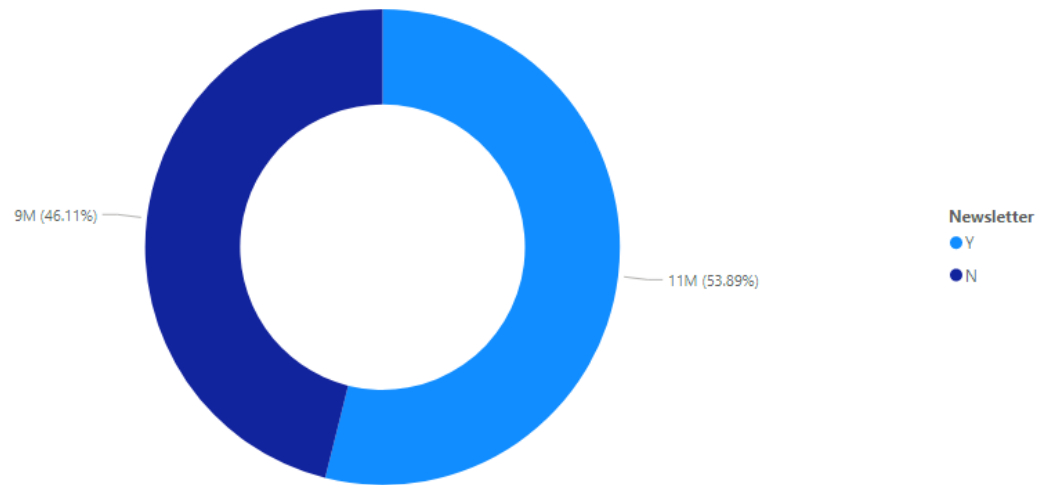
total_revenue by Country
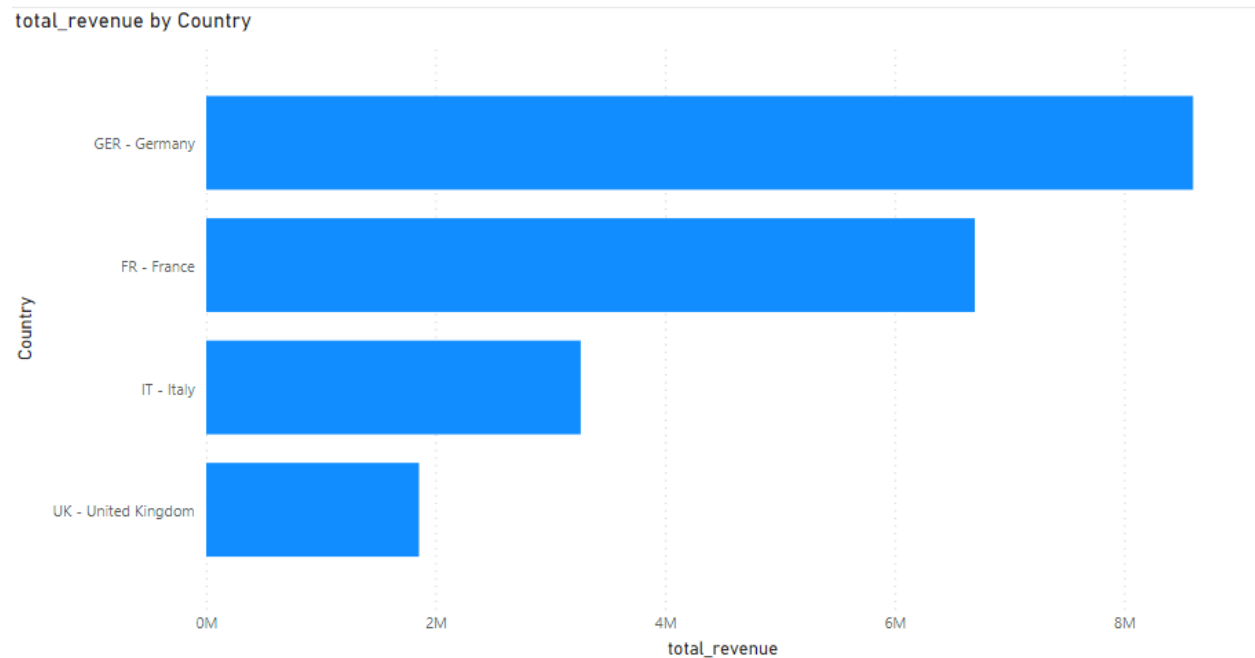
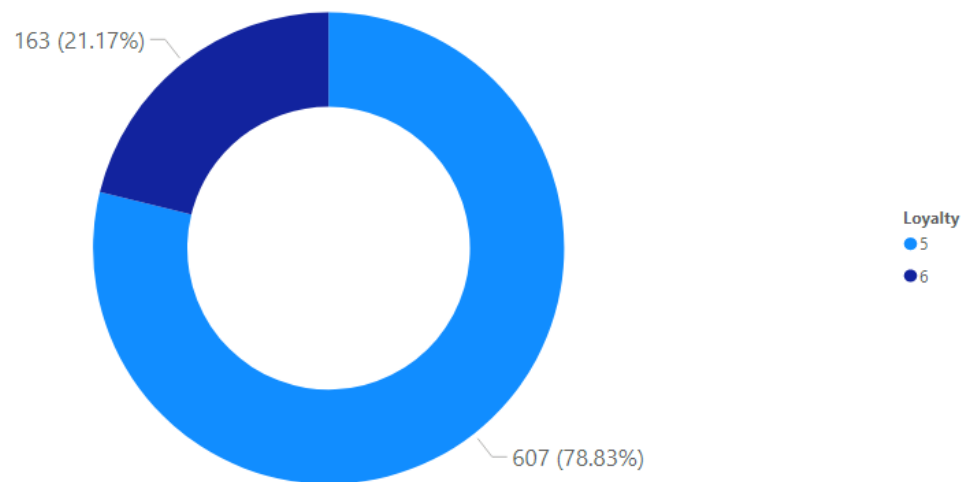## 4.2 Gender is not a factor when it comes to total revenue contribution

total_revenue by Gender
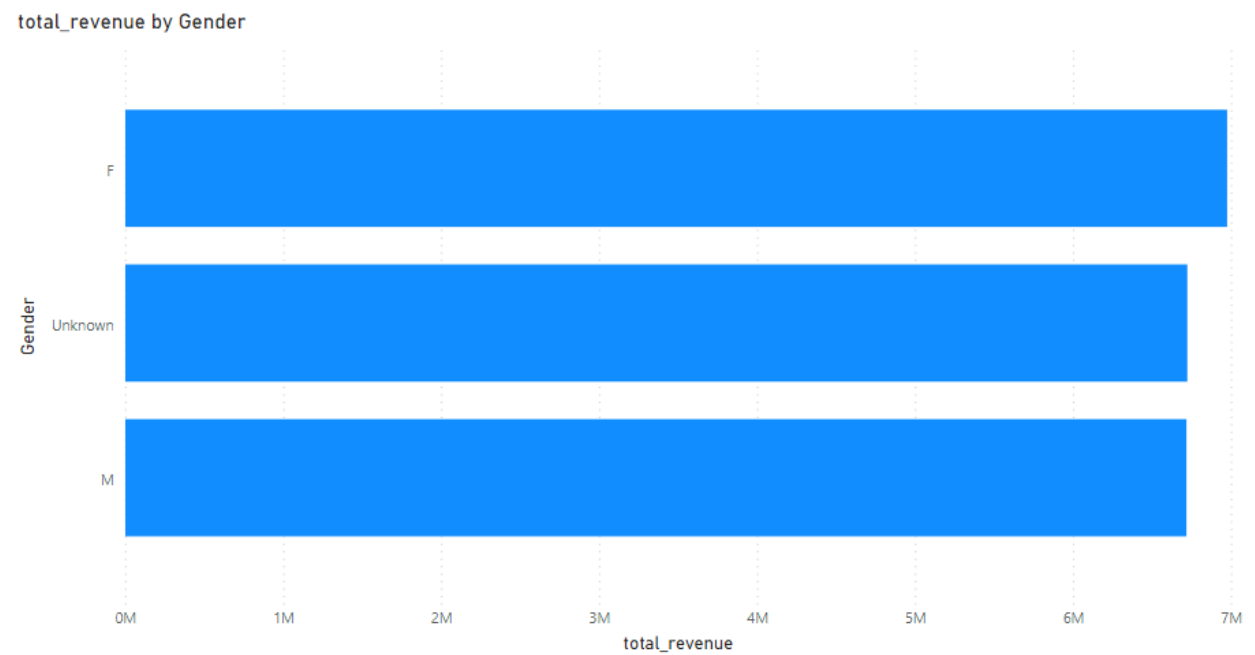
4.3 Newsletter subscribe customers contributes approximately 20% more revenue than who do not.

total_revenue by Newsletter

9M (46.11%)

11M (53.89%)

Newsletter
● Y
● N

4.4 Loyalty Age does not affect customer spending.

CustomerID by Loyalty


163 (21.17%)

607 (78.83%)

Loyalty
- 5
- 6

It seems that loyalty does not affect total revenue since the percentage of contribution in total revenue are also 20 – 80, which means that customers who have 6 years loyalty age don't spend more or less than those have 5 years.

total_revenue by Loyalty

4M (21.11%)

16M (78.89%)
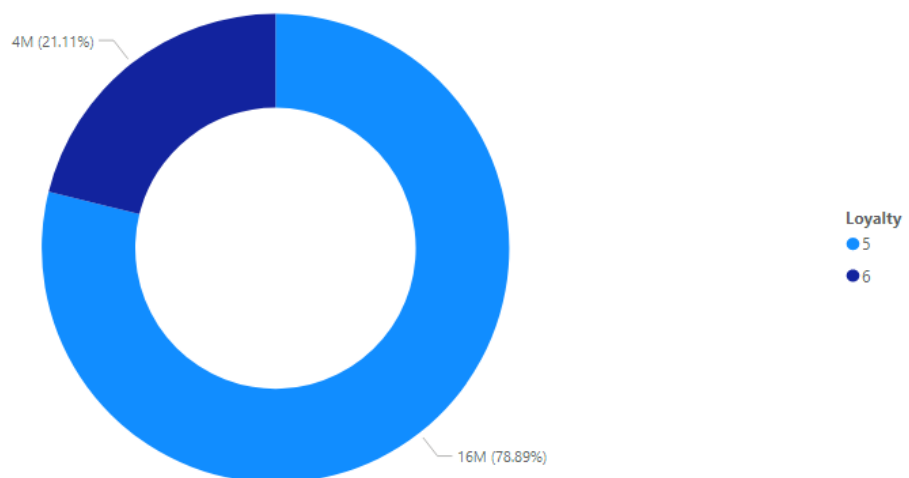
Loyalty
- 5
- 6

4.5 Age is not a factor when it comes to total revenue contribution

total_revenue by Age (bins)

3M (13.62%)

5M (24.96%)

4M (19.37%)

5M (22.22%)

4M (19.83%)

Age (bins)
- 30
- 20
- 40
- 10
- 50
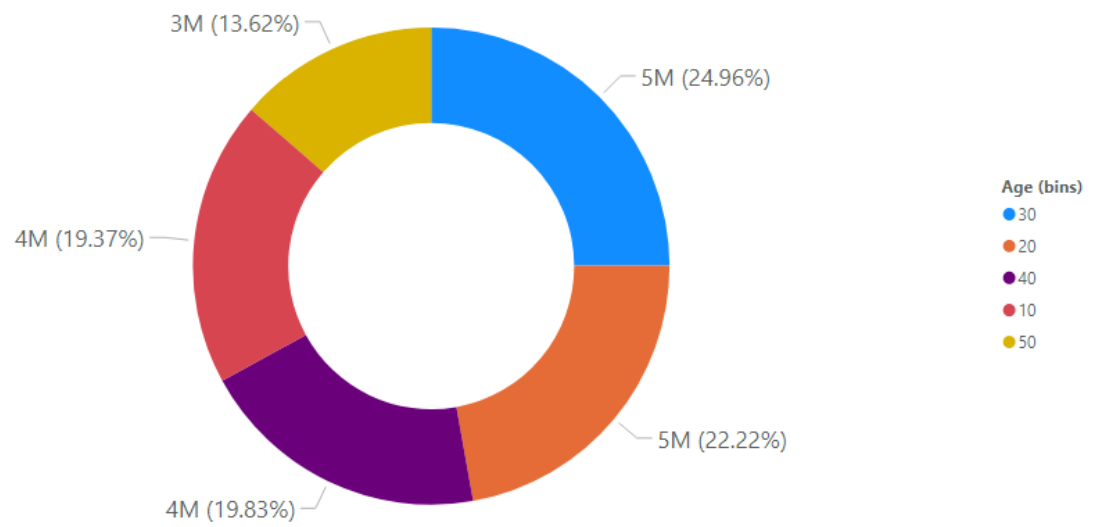
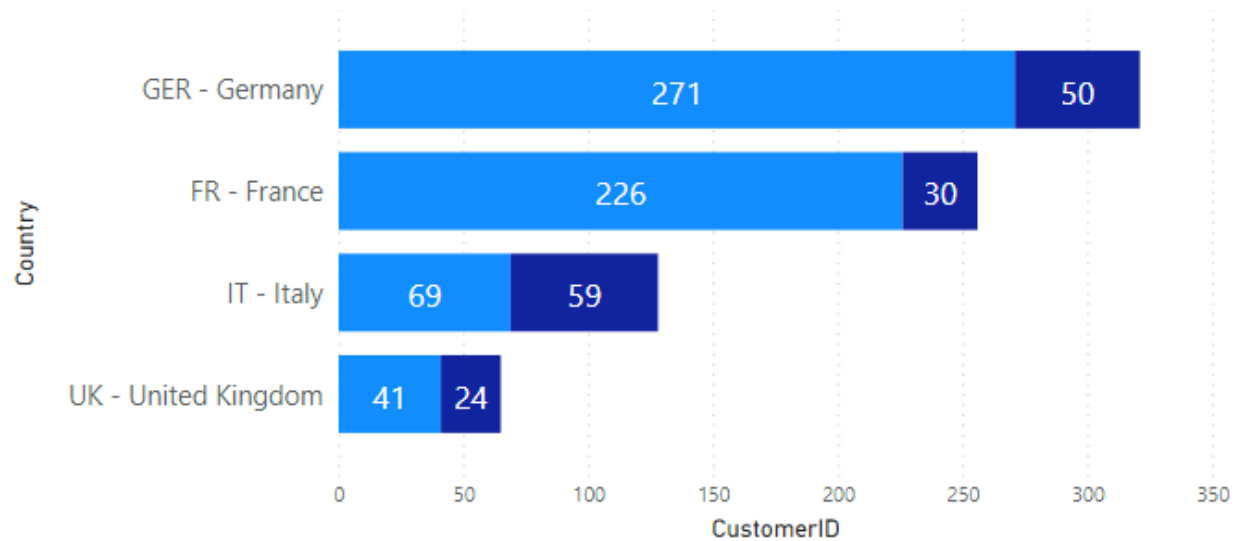## 4.6 Italy and the UK customer has the largest 6 year loyalty age percentage in total customers.

Although Italy and the UK people are the 2 smallest group, the have more than 30% of the total number customers are 6 years loyalty.
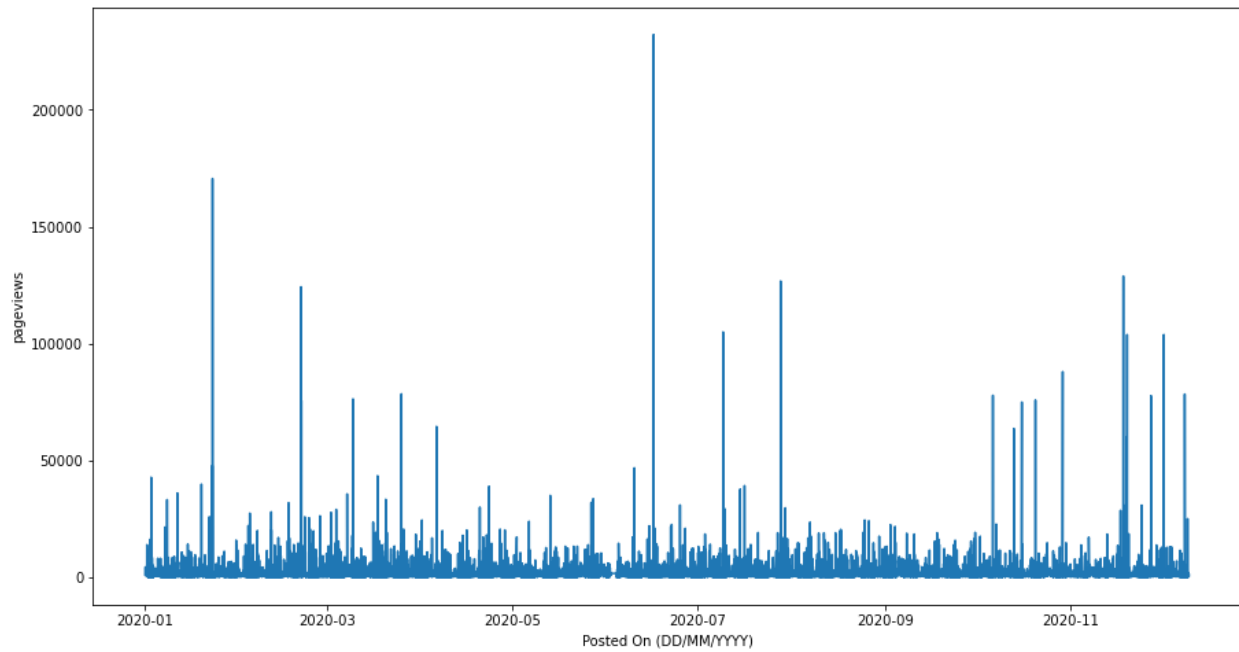
CustomerID by Country and Loyalty
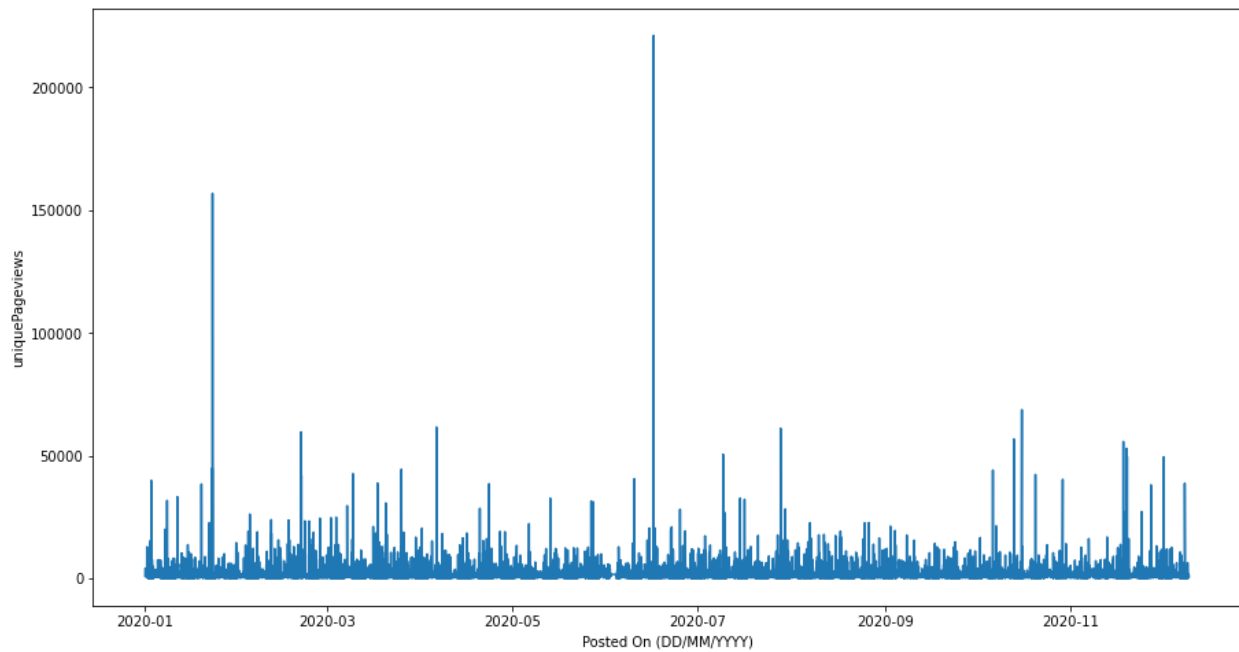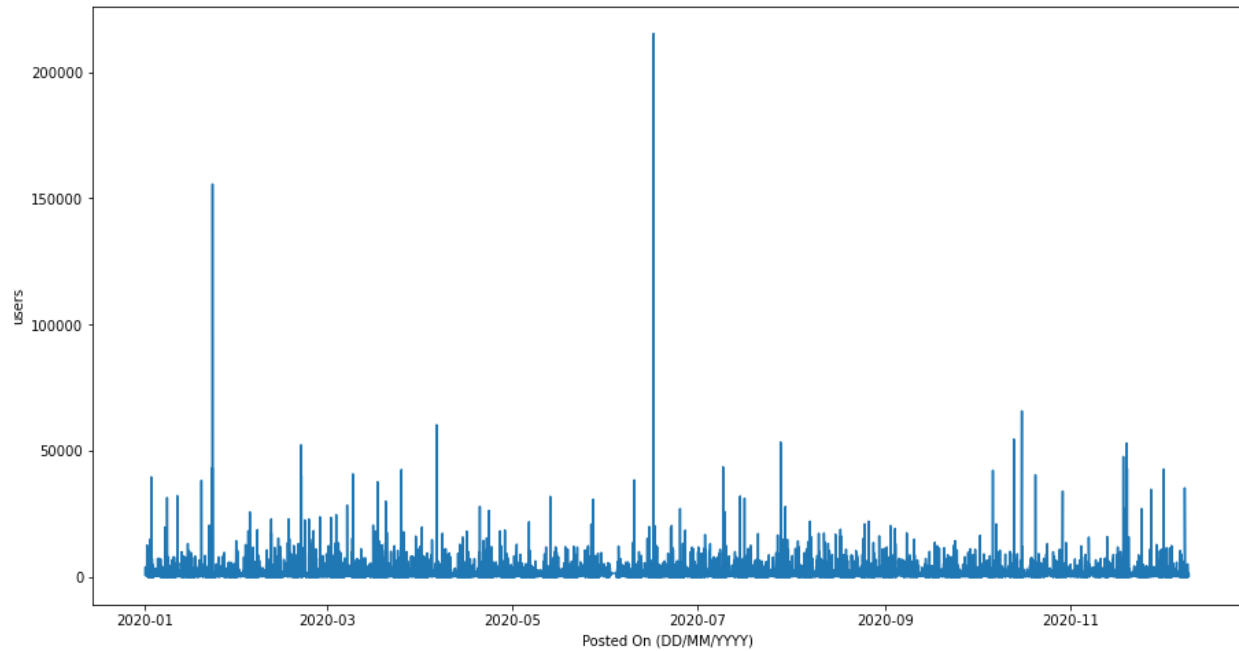
Loyalty ● 5 ● 6

Appendix

Comparision between pageviews, uniquepageviews and users in 2020

Pageviews in 2020

Uniquepageviews in 2020

Users in 2020

The revenue between In Store and Online is indifferent

	total_orders	total_customers	total_items
count	324.000000	324.000000	324.000000
mean	38.793210	38.793210	38.793210
std	6.310134	6.310134	6.310134
min	22.000000	22.000000	22.000000
25%	34.000000	34.000000	34.000000
50%	39.000000	39.000000	39.000000
75%	43.000000	43.000000	43.000000
max	58.000000	58.000000	58.000000

The statistics of total_orders , total_customers and total_items of Online Channel

	total_orders	total_customers	total_items
count	324.000000	324.000000	324.000000
mean	39.024691	39.024691	39.024691
std	6.549124	6.549124	6.549124
min	19.000000	19.000000	19.000000
25%	35.000000	35.000000	35.000000
50%	39.000000	39.000000	39.000000
75%	43.000000	43.000000	43.000000
max	60.000000	60.000000	60.000000

The statistics of total_orders , total_customers and total_items of In Store Channel

We can see that the daily average total_orders of both In Store and Online are nearly similar (38.79 compare to 39.02) , the ups and downs in total_orders between each day is also indifferent (standard deviation of 6.3 compare to 6.55) in the two channels.

	total_revenue
count	324.000000
mean	31465.910494
std	5616.831845
min	16626.000000
25%	27885.250000
50%	31282.500000
75%	35011.250000
max	47558.000000

	total_revenue
count	324.000000
mean	31550.194444
std	5748.147482
min	12005.000000
25%	27790.000000
50%	31760.000000
75%	35274.750000
max	51277.000000

Statistics of Total Revenue of Online Channel   Statistics of Total Revenue of In Store Channel

The average daily revenue between two channel is also approximately similar (31465 compare to 31550), the ups and downs in daily revenue between each day is also indifferent (standard deviation of 5616 compare to 5748) in the two channels.
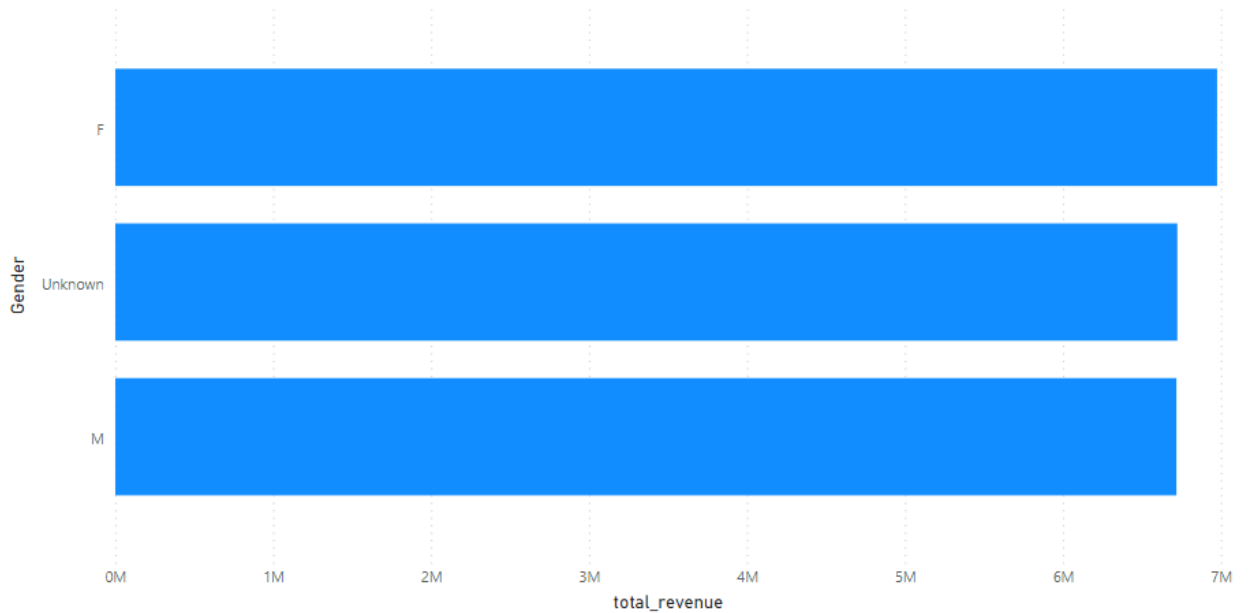
Customer Demographic
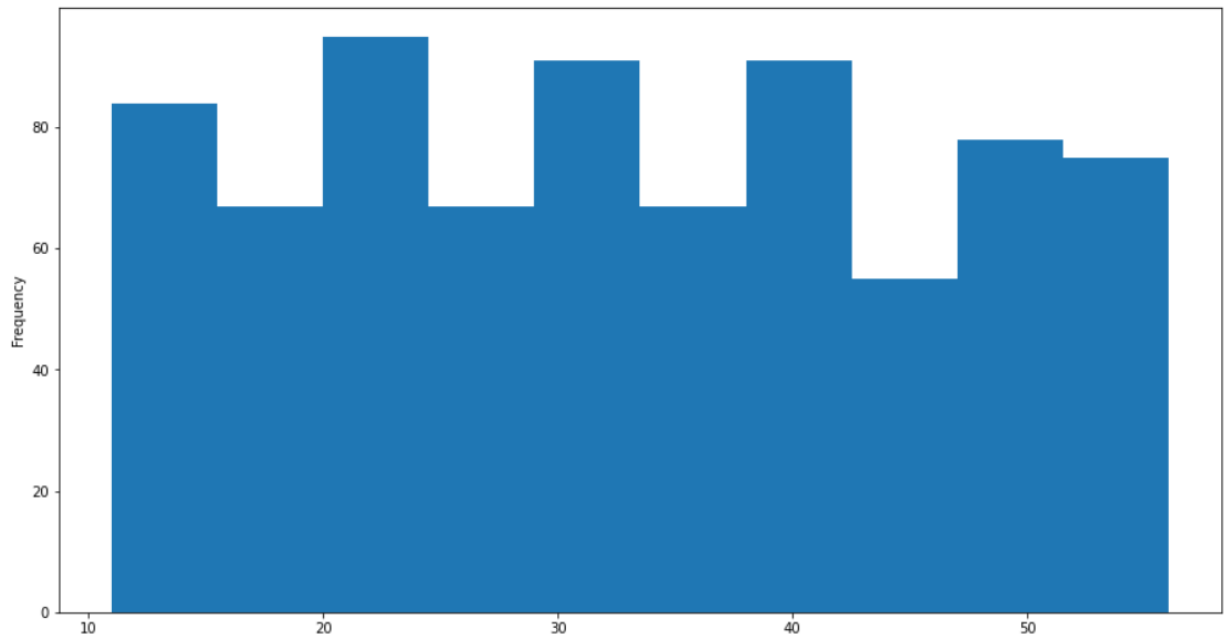
Age distribution of customers

It can be seen that customers varies from 11 years old to 56 years old, and they are distributed evenly into each groups ,there are no dominant age group in our dataset.

total_revenue by Gender

Notebook : Link